DOCUMENT RESUME

ED 174 677 TM 009 649

AUTHOR Crowder, Christopher R.
TITLE An Investigation of Item Bias Occurring at Different Ability Levels for Anglo and Mexican-American Students.
PUB DATE Apr 79
NOTE 31p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Anglo Americans; *Cultural Differences; Culture Free Tests; Failure Factors; Grade 7; *Item Analysis; Junior High Schools; *Mexican Americans; Reading Ability; Reading Level; *Reading Tests; Spanish Americans; *Test Bias; Vocabulary Development
IDENTIFIERS *California Achievement Tests; Distractors (Tests)

ABSTRACT

Two chi square techniques were applied to identify culturally biased items on the Reading Vocabulary and Reading Comprehension sections of the California Achievement Tests. Subjects were 1,035 Mexican Americans and 2,474 Anglo Americans in grade 7. Eight items appeared biased against Mexican Americans, but not Anglo Americans; seven occurred in the Vocabulary section. Three of the eight items also showed significant cultural group differences in response to incorrect answers, as well as four items not previously identified. Items were conditionally biased; that is, the bias appeared significant only for students in one level of reading ability--those answering 26-50% correct on the total test. Removing the biased items did not increase the mean score for the Mexican Americans. Mexican American educators cited the inability to understand stem words as the main source of bias for at least six of the items, rather than language usage, cultural values, or test wiseness. In conclusion, the two methods detected biased items, but did not necessarily identify the same items. Used jointly, they were unable to identify a biased item by achievement level, or to specify the incorrect response that was the source of bias. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED174677

Session No. 11.27

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Christopher R. Crowder
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

An Investigation of Item Bias Occurring at Different Ability Levels for Anglo and Mexican-American Students

Introduction

Previous methods for assessing test item bias have analyzed items with regard to total test score, item difficulty, item-total test correlation, item characteristic curve and distribution of incorrect responses. These methods have not offered an external criterion by which to judge the bias. One such method, introduced by Scheuneman (1975), employs a modified chi square to identify potentially biased items. With this approach, bias is defined as individuals within the same ability level, but of different ethnic backgrounds, having different probabilities of the correct response to an item. The ability of the student is operationally defined as the total performance on the test. Scheuneman has indicated that the method, which reports as cell frequencies the number of students obtaining a correct response, is conceptually and computationally simple, and appears effective in screening out biased items during test construction. However, the procedure offers little evidence as to the source of bias.

A method which does give clues about the source of bias compares whether attraction to incorrect responses, or foils, is homogeneous across cultural or ethnic groups ("cultural variation"). Where two or more groups are drawn toward different foils (pull index), the evidence for bias is strong (Veale and Foreman, 1976). Here, "cultural bias" refers to a distortion in p-values (item difficulties) for the groups studied. The chi square is computed in a two-way contingency table,

where rows correspond to ethnic or cultural groups and the columns correspond to the foils. If the statistic is significant, one concludes there is an association between the groups and the foils selected. Like the modified chi square technique, the method is conceptually simple and does not assume that the groups are representative of their respective populations.

Used together, the approach examining items by ability (or performance) level and the procedure comparing the pull of foils across groups offer two methods for detecting biased items that are relatively easy to calculate, mathematically, and can be used by school districts with comparatively small numbers of students in the ethnic groups to be studied.

In discussing bias, Green (1977) has noted that "the sensitivity of any test at its extreme is usually diminished," and that most groups of minority students do have low mean scores. Further, when such scores fall within a limited range, it indicates that only a subset of the items is functioning, and thus, the content validity is diminished. Roberts (1976) has stated that the level of test used is considered to be appropriate when the mean raw score is equal to or above one-third of the maximum and less than three-fourths of the total score, and that within this range the test should be suitable or at a functional level for the student tested. The same criterion can be applied to items of the test. Scheuneman's method of examining item bias by performance level enables one to study students scoring outside the functional range, by ethnic group, and compare their responses to those achieving within a functional range. Although the ability levels were previously

established separately for each item by Scheuneman to conform to the limitations of the chi square statistic (Siegel, 1956), the restrictions are not so great as to preclude an examination of the amount of item bias occurring within and outside the functional range of the test.

Once biased items have been identified by performance level and foil response patterns, the question as to the source of the bias becomes important to individuals or institutions trying to offset the effects of such bias. Basic sources of test bias have been identified (Green, 1977) as arising "from majority-minority differences in background knowledge, language, test wiseness and attitudes." All of these sources can be addressed by the schools, at least to some degree.

This particular study proposes to use both chi square techniques to help identify the biased item, the performance level of the student affected in terms of the test's functional level, the incorrect responses given for such an item, and the possible sources for the item's bias. More specifically, the following questions will be examined:

1. Can biased items and their distractors (foils) be identified on a reading comprehension test for Mexican-Americans and Anglos by ability level?

2. Will the items identified as biased show across all performance levels or indicate conditional bias for only certain performance levels?

3. Is conditional bias found primarily in the functional range, as identified by student performance level, or rather mainly outside the functional range of the test?

4. Which of the two subtests, Reading Vocabulary or Reading Comprehension, will show more biased items against Mexican-Americans and the majority group?

5. Will the removal of scores for items identified as biased increase the mean score for the ethnic group against which the bias occurred most frequently?

6. What sources of bias appear to be operating within the items identified--(a) background knowledge, (b) language usage, (c) cultural values or (d) test wiseness (knowing what is expected)?

Methods

All seventh grade students not in a self-contained learning disabilities class or who spoke only a language other than English were administered a norm-referenced reading test within a school district in the fall of 1978. The California Achievement Test, Reading Vocabulary and Reading Comprehension, Level 17, Form C, 1977 edition, was given to 1,035 students identified as Mexican-American (with Spanish surname) and 2,474 Anglos (the majority of district students) excluding American-Indian, Asian-American and Blacks.

The students were tested in their regular classrooms, and the answer sheets were machine scored. Each ethnic group, Mexican-American and Anglo, was then divided into four performance (ability) levels based upon the total reading score. The four performance levels were designated (a) one to 25 percent correct, (b) 26 to 50 percent correct, (c) 51 to 75 percent correct, and (d) 76 to 100 percent correct on the total test. The two middle levels or ranges (26 to 75 percent correct)

approximate what has been called the "functional" range of the test, and the lower and upper levels delimit ranges outside the functional range. The one to 25 percent correct level is comparable to a "guessing" range, since there are four distractors for most of the 70 items. The 76 to 100 percent correct level has been referred to (Roberts, 1976) as the "ceiling" range for the test.

Each item in the test was then analyzed using a contingency table with four performance levels representing one variable and the two ethnic groups the other variable. Scheuneman's modified chi squares were calculated, and those items with significant values at the .01 level of confidence were identified as having a possible bias. Computation of the magnitude of discrepancies between observed and expected values within each cell of the table for significant items revealed whether the bias crossed all four performance (ability) levels, or rather was "conditional," i.e., showed bias interacting with one or more performance ranges of the test.

A chi square was also computed for each item on the incorrect responses with the foils being one factor on the contingency table and the ethnic groups the other. Blank responses were included in the table with the foils to determine the probability for each incorrect response (pull index). Items with significant chi squares at the .01 level of confidence were identified as showing cultural variation and possible bias.

Thus, one method, Scheuneman's modified chi square, examined the correct answer to an item using total test score as an estimate of ability, whereas Veale and Foreman's chi square technique analyzed

incorrect responses to the item without assumptions regarding student ability. If the data permitted, where significant chi squares for foil responses were obtained on items, the response patterns were divided by performance level similar to that for the modified chi square technique, and the foil statistic rerun at each level. After the two chi square methods were run on each item, those items showing significant values ($p < .01$) by either method were analyzed with additional methods in an attempt to confirm the bias. The selected items were rank ordered by difficulty for each ethnic group (Angoff and Ford, 1973) and relatively large differences were noted. The item means, or p-values, were transformed as well into a normal deviate or z-value to obtain an interval scale, and the z-score transformed into a positive delta value using the equation $\Delta = 4z + 13$ (Angoff, 1972). For the two groups, the pair of delta values for the identified item was plotted on a graph and represented by a single point. The graph is equivalent to a correlation plot with a line of best fit. Relatively large deviations from this line represent items that are biased against one ethnic group. Also, point biserial correlations were computed between item responses and total score for each ethnic group. Values under .20 and differences in excess of .20 between the two groups were considered evidence of bias as stated by the test publisher (CTB/McGraw-Hill, 1978).

Items which indicated bias against Mexican-Americans by at least <u>two</u> of the statistical methods were given to a Mexican-American educator familiar with the culture to identify the possible sources of bias. The basic sources were identified as differences in (a) background knowledge, (b) language, (c) test wiseness and (d) attitudes and values (Green, 1977).

Further, some Mexican-American students who took the test and missed the biased items were questioned individually as to why they responded as they did.

The results which follow were gathered by the methods described in this section.

Results and Interpretation

Using Scheuneman's modified chi square method, eight items were identified as showing significant values ($p < .01$). Each of these items (6, 8, 12, 14, 16, 21, 29 and 52) appeared biased against Mexican-Americans at one or more ability levels (based upon total reading test performance). As shown in Table 1, seven of the eight items occurred in the Reading Vocabulary subtest rather than the Reading Comprehension subtest. No item appeared to be biased against Anglos, using this technique.

Since, with this statistic, the average proportion correct for an ability level is affected by extreme differences in ethnic group sizes, as indicated in Table 2, the proportion may not be representative of a particular group. However, when the size of the Mexican-American and Anglo samples were equalized by systematically sampling the larger majority group, no significant changes in chi square values were observed. It was concluded that the item bias identified against Mexican-Americans was not due to relative differences in the size of the majority and minority samples. A similar equating of samples on the other test items produced no new evidence of item bias against Anglos.

Table 1

Seventh Grade Reading
California Achievement Test, Level 17, Form C
Analysis of Items Biased Against
Mexican-Americans (MA) and Anglos (A)

Item Number	Item Difficulty Rank Order	$r_p$bis	Modified χ^2 by Ability Level	χ^2 for Incorrect Responses	Bias Against Ethnic Group(s)
Vocabulary 6	MA=41.5 A =19.5	.45 .54	33.58**	6.52	MA
Vocabulary 8	MA=16 A = 9	.44 .47	17.18**	19.76**	MA
Vocabulary 10	MA=55.5 A =67.5	.14 .35	6.25	20.45**	A
Vocabulary 12	MA=47.5 A =19.5	.29 .42	63.19**	6.66	MA
Vocabulary 14	MA=68 A =51.5	.31 .56	12.34*	41.82**	MA
Vocabulary 16	MA=51.5 A =29	.49 .56	16.59*	15.25*	MA
Vocabulary 21	MA=35.5 A =17	.54 .57	13.65*	5.30	MA
Vocabulary 29	MA=55.5 A =29	.49 .56	25.82**	1.35	MA
Comprehension 40	MA=41.5 A =58.5	.24 .27	1.77	20.02**	A
Comprehension 43	MA=44 A =55.5	.24 .37	1.28	36.68**	MA & A
Comprehension 50	MA=35.5 A =55.5	.20 .27	7.40	32.41**	A
Comprehension 52	MA=23 A =12.5	.46 .49	32.22**	6.70	MA

* $p < .01$
** $p < .001$

Table 2

Example of the Computation of Chi Square[a] for Item 6

Score on Total Test	Number with Scores in each Range			Obtained Frequencies (No. with item correct)			Proportion Correct	Expected Frequencies	
	Mexican-Americans	Anglos	Total	Mexican-Americans	Anglos	Total		Mexican-Americans	Anglos
	M	A	T	M_o	A_o	T_o	(T_o/T) p	(p·M) M_e	(p·A) A_e
54-70	61	778	839	58	760	818	.9750	59.47	758.55
36-53	223	907	1130	141	694	835	.7389	164.77	670.18
18-35	578	709	1287	139	286	425	.3302	190.86	234.11
1-17	173	80	253	36	27	63	.2490	43.08	19.92

$$\chi^2 = \sum \frac{(M_e - M_o)^2}{M_e} + \sum \frac{(A_e - A_o)^2}{A_e} = 33.58 \text{ d.f.} = 3$$

[a]Adapted from Scheuneman

When the item foils were tested across the two ethnic groups, three of the eight items (8, 14 and 16) identified by the modified chi square as being biased against Mexican-Americans, also showed significant differences in response to incorrect answers. Also, four items (10, 40, 43 and 50) not previously identified, demonstrated significant foil differences between Mexican-Americans and Anglos. An effort to discover whether significant foil differences occurred only between specific performance levels for all items showing foil differences was not successful. This was because, by ability level, more than 20 percent of the cells in the chi square contingency table contained expected frequencies less than five or at least one cell had an expected frequency less than one (Siegel, 1956). Thus, for almost all items identified as biased by a comparison of foils, the performance levels were combined (one to 100 percent correct).

What becomes apparent in comparing the modified chi square technique for ability (performance) levels and the chi square method for foil differences is that the two methods do not necessarily identify the same item as showing possible bias. Veale (1977) has noted that when the chi square method uses only correct response data, the statistical value produced is not invariant with respect to permutations of rows and columns. Thus, the modified chi square could yield different results depending upon whether the ethnic groups and ability levels were placed in the row or column margins. Also, with respect to foil analysis, two or more ethnic groups may respond to foils in a similar pattern, but item bias might still be detected in either the correct response choice,

the item body or the instructions for the item. Such bias could be identified through rank order differences if the item appears biased across all performance levels.

Additional statistical analyses were made on the items identified as revealing significant values by either the modified chi square by ability level or the chi square on foil responses. As indicated in Table 1, relative differences in rank order difficulty favored Anglos on all eight of the items recognized as biased against Mexican-Americans through the modified chi square method. The four items (10, 40, 43 and 50) identified through the chi square for the foil responses each showed that the relative rank order difficulty was greater for Anglos than for Mexican-Americans. Another statistic dealing with item difficulty, the delta value (Angoff, 1972), was calculated for each biased item for both ethnic groups, and the pairs of values per item plotted on a graph (see Figure 1). The regression line of best fit is plotted for the delta values, and a confidence interval, represented by two dotted lines, is calculated to indicate the probability ($p < .05$) of large differences in item deltas between Anglos and Mexican-Americans. Four of the items calculated to be biased against Mexican-Americans fall on or below the lower dotted line (Items 6, 12, 16 and 29), as shown in Figure 1, supporting this direction of bias. Further, each of these items had relative rank order differences between groups greater than 20. No outliers of this magnitude were identified on or above the upper dotted line or showed bias against Anglos.

Point biserial correlations for the eight items thought to be biased against Mexican-Americans ranged from a low of .29 to a high of .54,

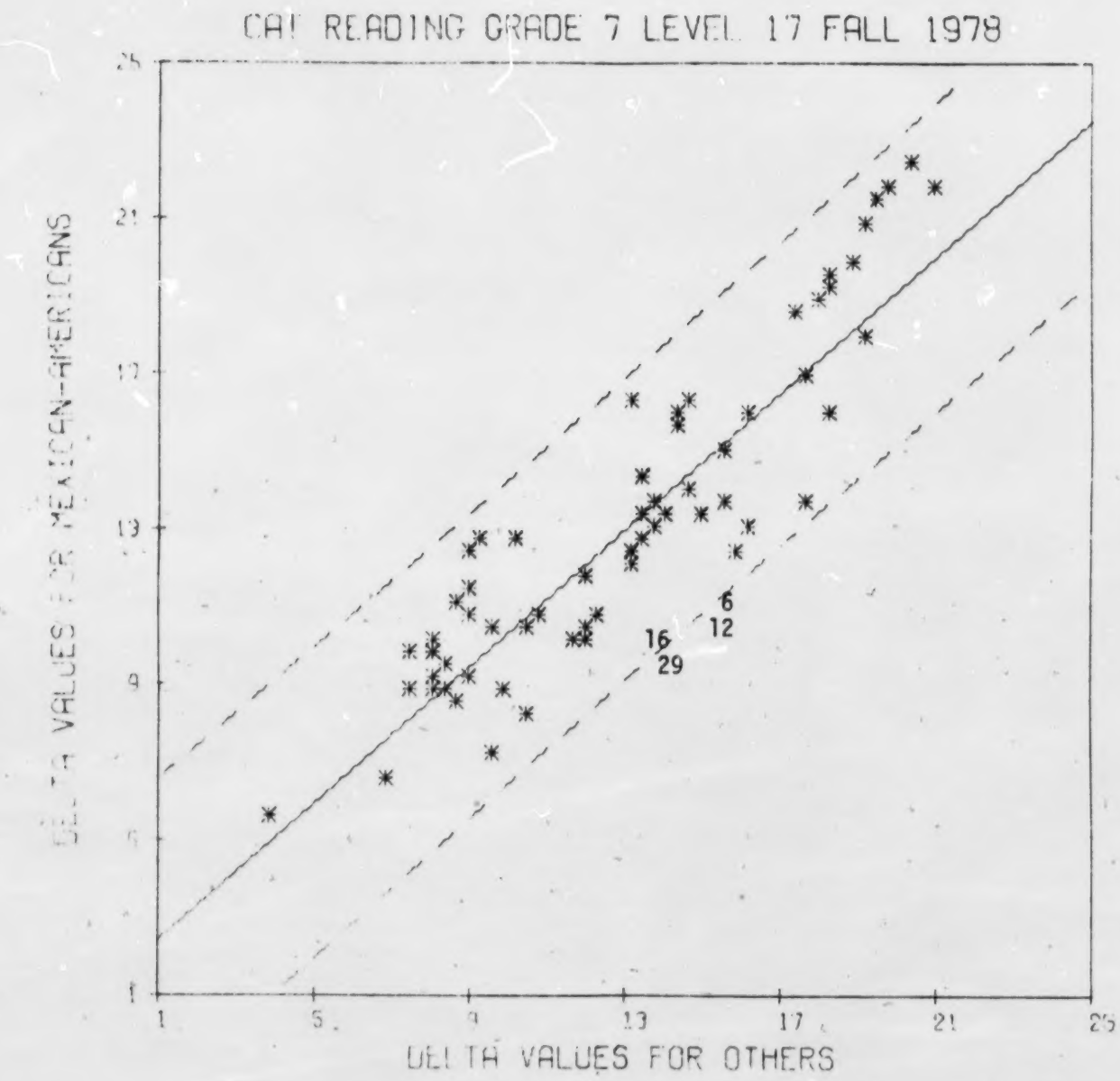

Figure 1. A bivariate plot of delta values for items on the seventh grade reading test for Mexican-Americans and Others (Anglos).

compared to a range of .42 to .57 for Anglos on the same items, as shown in Table 1. All of the items biased against Mexican-Americans except one (Item 14) had point biserial differences between the majority and minority group of less than .20. This difference, along with an item-total test correlation greater than .20, was established by the test publisher as minimal criteria in selecting items for the test (CTB/McGraw-Hill, 1978). Point biserial differences for the four items (10, 40, 43 and 50) favoring Mexican-Americans by the rank order method ranged from a low of .14 to a high of .24 for Mexican-Americans and from .27 to .35 for Anglos. Both ranges represent relatively low item-total test correlations, and one, Item 10, would have been discarded by the publisher's own criterion if the district's Mexican-Americans had represented the sample for the minority on item tryout or standardization. The test publishers (CTB/McGraw-Hill, 1978) have indicated, however, that Hispanic item data were not examined separately, in contrast to that for Blacks, in the development of the items for level 17 of the test.

Each of the eight items classified as biased against Mexican-Americans by the modified chi square technique was identified as being "conditionally" biased, i.e., the bias appears significant only in specific performance (or ability) levels. The conditional bias can occur in one or more of the four levels (based upon total test score), and two of these categories, the highest and the lowest, fall outside the approximate functional range of the test (the two middle levels). None of the eight items manifested conditional bias by performance level <u>outside</u> the functional range. However, one item (12) revealed

a slight trend toward bias in both the one to 25 percent correct and 76 to 100 percent correct performance levels by differences in the percent correct between the majority-minority groups at these levels. One intervening variable in such an analysis is that the number of subjects scoring at the extreme levels is generally smaller than the number scoring in the functional range for the test. Since the modified chi square is affected by the size of the sample, it is not clear whether the apparent differences occurring within and outside the functional range are real, or an artifact of the method. Scheuneman (1977) has indicated that a minimum of 10 to 20 observed correct responses per cell produces more stable outcomes in predicting item bias, but the problem should be explored further.

Considering such limitations, the average differences in the percent correct between Mexican-Americans and Anglos by performance level is, nevertheless, greater in the approximate functional range than outside the functional range for items identified as biased against Mexican-Americans. Table 3 demonstrates this with the differences between the two ethnic groups and the average difference (with same number of subjects per item) in each of the four performance ranges.

The four items identified as biased by the chi square on foil responses (Items 10, 40, 43 and 50) did not show significant modified chi square differences by performance level, and the question of whether the item bias occurred primarily within or outside the functional range could not properly be addressed. An examination of the average difference in percent correct between the majority-minority groups by performance level showed no apparent differences between the four levels.

Table 3

Differences Between Mexican-Americans and Anglos in Percent Correct for Biased Items by Four Performance Levels

	Percent Difference* by Performance Level on Total Test			
Item	1-25%	26-50%	51-75%	76-100%
6	-10	-14	-16	-2
8	- 5	-17	- 8	-2
12	- 7	-22	-28	-6
14	- 1	- 2	-20	-4
16	- 2	-10	-15	+3
21	- 7	-12	- 7	0
29	1	-15	- 6	-3
52	-10	-18	-11	-1
Average	- 5.1	-13.8	-13.9	-1.9

*Positive differences in percent correct favor Mexican-Americans, and negative differences favor Anglos

Differences within each performance level, however, generally favored Mexican-Americans on three of the four items (10, 40 and 50).

In an attempt to determine what effect the removal of items identified as being biased would have on the total score for the group, the eight items thought to be biased against Mexican-Americans were removed from the test. A similar but separate procedure was conducted for the three items appearing to be biased against Anglos (Item 10,

40 and 50). Table 4 demonstrates the effect of this procedure. Compared to the unadjusted total average percent of items correct, the separate removal of biased items for either ethnic group produces changes of less than one percent in each group. The difference in percent correct for the majority-minority group changes less than two percent with either adjustment. Thus, although there are eight items showing conditional bias against Mexican-Americans, the removal of these items reveals little change in the overall average percent correct of the remaining items nor any pronounced change in the relative average percent correct between the Mexican-Americans and Anglos.

Table 4

Effects of Removal of
Biased Items Separately Against
Mexican-Americans and Anglos

Biased Items Removed	Average Percent Correct		
	Anglos	Mexican-Americans	Difference
for:			
Mexican-Americans	61.6%	42.3%	19.3%
Anglos	63.2%	42.0%	21.2%
Unadjusted	62.5%	41.7%	20.8%

This relatively small change may be due, at least in part, to the fact that the bias identified was conditional and did not show across all performance levels for each item. Table 4 shows little change in average

percent correct either when the three items biased against Anglos are discarded. Since there are fewer items biased against Anglos, such results might be more expected than for the eight items biased against Mexican-Americans.

In an effort to discover what are some of the basic sources of bias for the items previously identified, a Mexican-American educator knowledgeable about cultural variations within the minority community tested was shown the identified items and the differences in response foil patterns. Also, a limited sample of Mexican-American seventh grade students who missed the biased items were queried by their teacher as to why they responded as they did.

The Appendix identifies the chi square foil response patterns for each identified biased item and reports the combined input from the Mexican-American educator and students. Of the eight items thought to be biased against Mexican-Americans, the source of bias for at least six of these items seems primarily due to differences in background knowledge. More specifically, the Mexican-American students lack the vocabulary to determine what the stem word means, rather than what the response choices mean. The students reported that they used several incorrect heuristic devices trying to figure out an unknown word, but few contextual clues, if any, are offered for such vocabulary words. For one item (21), Mexican-Americans seemed to have difficulty in following the directions and shifting to a new task, that of choosing word opposites, in the middle of the Reading Vocabulary subtest. Developing the minority students' test-taking skills might correct this difficulty. One additional item (52) on the Reading Comprehension subtest was apparently missed because

Mexican-American students were not familiar with similes, and did not key on the meaning of the word "like" to answer the question.

The three items (10, 40 and 50) apparently biased against Anglos all show a pattern of students being significantly attracted to a likely foil, based upon partial knowledge and comprehension of the word (Item 10), poem (Item 40) or informational paragraph (Item 50). The response patterns to the foils are somewhat more evenly distributed for the Mexican-Americans on each of these items, indicating that a lack of background knowledge may not make likely foils as appealing to this group.

Finally, one item (43) contains a response so attractive to both ethnic groups, that it is chosen more frequently than the correct response, a condition given by the test publisher (CTB/McGraw-Hill, 1978) for discarding the item. Students in this community appear to let an attitude toward a clean environment overly influence their response to the item.

Conclusions

The evidence for this study indicates that biased items can be identified through the use of the modified chi square technique by ability levels or the chi square method of foil analysis. However, the two methods do not necessarily identify the same items as being biased. The hope (perhaps naive) that the two methods, used jointly, would identify a biased item by ability or performance level and specify the foil within the recognized level that is the source of bias, was not substantiated by this study. The size of each group studied, the variance possible in the modified chi square value, the limitations of the type of bias identifiable with foil analysis, and the difficulty of the test, are all factors

which could produce differences between the two methods.

The items shown to be biased by the modified chi square method, all revealed evidence of conditional bias, that is, the differences between ethnic groups interacted with one or more (but not all) performance levels. The question remains as to why no biased items across ability levels were discovered. This result may be at least partially due to the limited number of students generally found in the highest and lowest performance ranges, and the related effect on the chi square statistic.

The conditional bias discovered in the items occurred only in the more functional range of the test, i.e., in the 26 to 75 percent correct performance level. One might conclude that bias at this level can be identified, because the test is more sensitive to differences in instruction, knowledge, or other sources compared to the approximate "guessing" and "ceiling" levels of the test. However, since the number of subjects scoring at the two extreme levels is usually much smaller than in the functional range, the results may be more an artifact of the method than real.

The use of additional statistical procedures, especially those comparing differences in item difficulties between the groups, supported items previously identified as being biased, both against Mexican-Americans and against Anglos. The computation of relative rank order differences and/or delta values seem to be important additions in confirming suspected bias. It is highly recommended that such statistics be used by school districts to screen all items for possible bias, and then to do further analysis, such as chi square on foil responses, for items with relatively large rank order differences.

In contrast, the point biserial statistic did not appear to be that sensitive in identifying ethnic bias within particular performance groups, especially for items whose rank orders occur in the middle range. Nevertheless, the statistic is useful in confirming hard items that fail to discriminate across performance levels for one or more ethnic groups.

The removal of scores for items recognized as biased against Mexican-Americans and Anglos separately, showed only slight changes in mean percent correct compared to the unadjusted means for each ethnic group. The fact that all items identified as biased against Mexican-Americans were conditional, and the bias did not extend across all levels, may account for such slight changes. For Anglos, only a few items were shown to be biased. That such differences were small would preclude eliminating this test from a district testing schedule. It is probable that few other standardized tests would fare much better in such a study. Fewer items biased against Mexican-Americans might have been identified, if the item tryouts had included separate data for Hispanics, as they did for Blacks. In a separate study, fewer biased items were found for Blacks than for Mexican-Americans in the same district and grade.

The attempt to recognize some of the sources of bias, using feedback from an educator knowledgeable about Mexican-American cultural variations and input from a few minority students who missed the item, indicated that the main source of bias was due to differences in background knowledge. That the source appeared mainly in the Reading Vocabulary subtest is not surprising, since few if any contextual clues for the stem word of the item are offered here to the student. Some

responses from Mexican-American students revealed that even without such clues, they tried different methods in arriving at an answer, rather than guessing randomly. The implications are that the minority students need instruction in recognizing English vocabulary, presented in isolation.

The source of bias for items against Anglos was identified through foil differences, with the majority student who has partial knowledge being more frequently attracted to the most difficult foil than the minority student. Differences in prior instruction between the groups might also account for such variations.

To determine if the results reported in this study are reliable, the study should be replicated with seventh grade students from the same district in the coming school year. Also, an analysis by sex and age, as well, might shed more light on the possible sources of item bias.

References

Angoff, W.H. A technique for the investigation of cultural differences. Paper presented at the annual meeting of the American Psychological Association, Honolulu, May 1972.

Angoff, W.H. & Ford, S.F. Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 1973, 10, 95-105.

CTB/McGraw-Hill. California Achievement Test Technical Bulletin 1. Monterey, California, 1978.

Green, D.R. Problems of achievement tests in program evaluation. In M. Wargo & D. Green (Eds.), Achievement testing of disadvantaged and minority students for educational program evaluation. California: CTB/McGraw-Hill, 1977.

Roberts, A.O.H. Out-of-level testing. ESEA Title I Evaluation and Reporting System (Technical Paper No. 6). Mountain View, California: RMC Research Corporation, October 1976.

Scheuneman, J. A new method of assessing bias in test items. Paper presented at the annual meeting of the American Educational Research Association, Washington, D.C., April 1975.

Scheuneman, J. A method of assessing bias in test items in the absence of an outside criterion. Preliminary draft, The Psychological Corporation, 1977.

Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.

Veale, J.R., A note on the chi square test with "correct/incorrect" data to detect culturally biased test items (S'(Technical Report No. 4). Berkeley, California: Statistical Research: Education and the Behavioral Sciences, December 1977.

Veale, J.R. & Foreman, D.I. Cultural variation in items and tests: A "global" analysis (REAL Technical Report No. 1). Denver, Colorado: Council for REAL Education, 1976.

APPENDIX

Appendix

Differences in Foil Response Patterns and Possible Sources of Bias for Items Identified Against Mexican-Americans and Anglos.

Item	Foil Response Pattern	Possible Source of Bias
No. 6	No significant differences in foil patterns are observed, and both ethnic groups choose "challenge" most frequently as a synonym for the word "baffle."	Mexican-American students who missed the item, stated that they didn't know what the stem word meant. Also, this word is first in the phrase, and few contextual clues for the meaning of the word are present. The source of bias appears due to differences in background knowledge.
No. 8	Both ethnic groups choose the foil "eliminate" most frequently as a synonym for the word "detect." However, Mexican-Americans are pulled more to both the foils "dislike" and "increase" than Anglos.	Again, Mexican-American students who missed the item indicated they did not know the meaning of the stem word. The source of bias appears likely to be differences in background knowledge.
No. 10	Anglos choose the foil "factor" most often as a synonym for the word "trait," whereas Mexican-Americans select the foil "ambition" most frequently.	One foil appears to lower the item success of the Anglos. It is frequently used in the English language, as for example, "important factor," and may be less well-known to Mexican-Americans. The source of bias may be due to background knowledge or language usage.

Appendix

Differences in Foil Response Patterns and Possible Sources of Bias for Items Identified Against Mexican-Americans and Anglos.

Item	Foil Response Pattern	Possible Source of Bias
No. 12	No significant differences are observed in the foil patterns between the two groups, and both pick the foil "fall continually" for the word "smolder," although Mexican-Americans often choose the foil "grow rapidly."	Mexican-American students indicate that they did not know the stem word, and there is no Spanish equivalent for the English term. The source of bias seems primarily due to differences in background knowledge.
No. 14	Foil response differences are significant with both groups choosing the foil "dangerous" most frequently, but Mexican-Americans are attracted more to this foil, whereas Anglos often select "covered" for the stem "massive."	The source of bias against Mexican-Americans is likely due to background knowledge, since some who missed item said they did not know the stem word, but it "sounded" like "dangerous."
No. 16	Mexican-Americans left the item blank significantly more than did Anglos.	Mexican-Americans who missed the item sometimes used a close cognate, "descripcion," to select the foil "picture of an artist" as an answer, but generally indicated they didn't know meaning of stem word "inscription."

Appendix

Differences in Foil Response Patterns and Possible Sources of Bias for Items Identified Against Mexican-Americans and Anglos.

Item	Foil Response Pattern	Possible Source of Bias
No. 21	No significant foil differences are observed between the ethnic groups, and both choose the foil "break" most often as an antonym for the word "depart."	This is the first item of five in which the students are asked to choose the word opposite, and some Mexican-Americans indicated that they did not understand the task, initially. The Spanish "partir" can mean to break apart, the most frequently chosen foil. The source of bias appears at least partially due to lack of test-taking skills.
No. 29	There are no significant differences in the foil response patterns between the two groups: Here students are required to select the sentence in which the underlined word means the same as stem word(s).	Some Mexican-Americans indicated that they only knew what the verb "supports" meant, but not the definition for the noun form of the stem word. The source of bias seems mainly due to dissimilarity in background knowledge.
No. 40	Anglos choose the foil "describes the soothing sound of the poem" most frequently for the meaning of the word "lullaby" in the context of the poem. Mexican-Americans do not select this foil as often.	Anglos either fail to read the poem or do not understand it and are drawn to a foil, which commonly describes the word "lullaby." Mexican-Americans may not be as familiar with the English meaning. The source of bias is not clear, but may be mainly due to differences in background knowledge.

Appendix

Differences in Foil Response Patterns and Possible Sources of Bias for Items Identified Against Mexican-Americans and Anglos.

Item	Foil Response Pattern	Possible Source of Bias
No. 43	Both ethnic groups choose "encourage listeners to clean up the environment" most frequently as a foil for what the speaker in a passage is trying to accomplish, but, Anglos select the foil more than Mexican-Americans.	This item appears to be a poor item in that an incorrect response is chosen more often than the correct answer by both ethnic groups. The source of bias may be due to differences in this region in cultural values, since several students in both groups felt that "cleaning up the environment" was most important.
No. 50	Anglos choose the foil "number of silver particles" significantly more than Mexican-Americans, who are more evenly divided on their incorrect responses.	The foil pulling Anglos is in close approximation to the correct answer, and although the item is difficult for the group, their partial knowledge of developing photographs seems to attract them to an incorrect response.
No. 52	The differences in foil responses between the two ethnic groups is not significant, but Mexican-Americans tend to choose the foil "dangerous" more than do Anglos for the meaning of the phrase "lazy snakes."	Several Mexican-Americans who missed this item thought that a description of back roads "that stretched out like lazy snakes over the hills" literally meant that there were snakes on the road, which made it "dangerous." Thus, the minority group tends not to recognize a simile, and the source of bias seems mainly due to differences in background knowledge and usage of the language.